Assignment 0

1. Vectors: Define the scalar product of two vectors.

- Scalar product of 2 vectors can be defined as the product of the magnitude of both the vectors and the cosine of the angle between them. It is also known as the dot product of the vectors.
- It is represented as:
  $A.B = |A||B|cos(\theta)$

2. Matrices (1): Define the product of two n x n matrices A and B and the transpose of a matrix

- The product of 2 n*n matrices will result into a single n*n matrix.
- The elements of the product matrix are the sum of multiplying each row of matrix A to the corresponding element in each column of matrix B.
- $C = AB = [c_{ij}]$

  $c_{ij} = \sum_{k=1}^{n} a_{ik} * b_{kj}$
- The transpose of a matrix is obtained by swapping the rows of the matrix with its columns. It is denoted as A^T, transpose of matrix A.
- If $A = [a_{ij}]$, then the elements of the transpose matrix $A^T = [b_{ij}]$ are given by:

  $b_{ij} = a_{ji}$

3. Matrices (2): What are the eigenvectors and eigenvalues of a matrix?

- For a square matrix A, a scalar λ is called an eigenvalue if there exists a non-zero vector v (the eigenvector) such that: $Av = \lambda v$
- Therefore, we can also say that when the matrix A is multiplied by its eigenvector v, the result is a scalar multiple of v. Here, λ is the eigenvalue corresponding to v.
- An eigenvector v associated with an eigenvalue λ is a non-zero vector that satisfies the equation $Av = \lambda v$

4. Statistics (1): What is the definition of mean, median, and standard deviation?

- Mean and median are the measure of central tendencies of the data.
- Mean/average is obtained by dividing the sum of all the values by the number of values present.

  $\mu = \frac{\sum_{i=1}^{n} x_i}{n}$, where μ is the mean of the n values in the data
- Median of the data is the middle or the central value in the sorted data. Depending whether the number of observations in the data are even or odd, the median is defined as follows:

  With odd number of elements, the median is at position $\frac{n+1}{2}$

  With even number of elements, the median is the average of the values at positions $\frac{n}{2}$ and $\frac{n}{2} + 1$

- Standard deviation measures the spread or the dispersion of the dataset. It represents the variation of the data points from the mean.
- $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$, where σ is the standard deviation and μ is the mean.

5. Statistics (2): What is Bayes theorem? What is the likelihood?

- Bayes' Theorem defines the probability of an event based on prior knowledge of conditions that might be related to the event.
  For 2 events A and B, Bayes' theorem is given as:
  $P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$
  Here, P(A|B) and P(B|A) are the conditional probabilities.
- The term "likelihood" refers to the probability of observing a set of data given a particular hypothesis or model. Likelihood plays a crucial role in Bayes's Theorem. The numerator P(B|A)*P(A) can be viewed as the likelihood of B given A, multiplied by the prior probability of the hypothesis.

6. Statistics (3): Define a discrete probability distribution pk (k=1, 2, 3, ...). Define the k-th moment of the distribution. Define the Bernoulli and the Poisson distribution.

- A discrete probability distribution is a statistical distribution that describes the probabilities of the possible outcomes of a discrete random variable. It associates each potential value of the random variable with a probability.
- The k-th moment of a probability distribution provides a measure of the spread of the distribution. It involves multiplying each possible value of the random variable by its probability and summing up these products.
- The Bernoulli distribution is a discrete probability distribution with two possible outcomes: success(1) and failure(0). It is represented by a single parameter, *p*, which is the probability of success.
- The Poisson distribution models the number of events occurring in a fixed interval of time or space. It is represented by *λ*, which means the average rate of occurrence of the events.

7. Statistics (4): Consider the power law distribution: p(x) = C x-4 , defined for x ≥ 1. What is the value of C?

- Here, C is the constant when the total probability is 1.
  $\int_1^\infty Cx^{-4}dx$ = 1 this is the pdf over the function p(x)
  = $\lim_{a\to\infty} \int_1^\infty Cx^{-4}dx$, now we solve for
  $\lim_{a\to\infty} [-\frac{C}{3}a^{-3} + \frac{C}{3}] = 1$
  $\lim_{a\to\infty} [-\frac{C}{3a^3} + \frac{C}{3}] = 1$, here the first term will tend to 0.
  $\lim_{a\to\infty} [\frac{C}{3}] = 1$

$\frac{C}{3} = 1$ and thus, $C = 3$

8. Statistics (5): define the generating function of a discrete probability distribution pk (k=1, 2, 3, ...).

- The Probability Generating Function is useful for calculating probabilities associated with the distribution.

  $G(z) = \sum_{k=0}^{\infty} pk\, z^k$

- The moment generating function is useful for calculating moments of the distribution.

  $M(t) = \sum_{k=0}^{\infty} pk\, e^{tk}$

9. Statistics (6): Given two variables X and Y, both with n entries (x1 , x2 , ..., xn and y1 , y2 , ..., yn), define the Pearson correlation coefficient between them.

- The Pearson correlation coefficient is a measure of the linear relationship between two variables. The formula for calculating the Pearson correlation coefficient between two variables X and Y:

  $$r = \frac{\sum_{i=1}^{n}(xi - \overline{X})(yi - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(xi - \overline{X})^2 \sum_{i=1}^{n}(yi - \overline{Y})^2}}$$

  Here, xi and yi are the data points

  $\overline{X}$ and $\overline{Y}$ are the respective means of the datasets.

- If,

  r = 1, then positive correlation

  r = -1, then negative correlation

  r = 0, then no linear correlation

10.

- The differential equation is given by:

  $\frac{dS}{dt} = -\beta S I \frac{dI}{dt}$

- Here,

  S: Represents the fraction of the population that is susceptible to the infectious disease. Individuals who are not infected but can become infected.

  I: Represents the fraction of the population that is infected with the disease.

  β: Is a positive constant representing the transmission rate of the disease.

  dS/dt : Represents the rate of change of the susceptible fraction with respect to time.

  dI/dt : Represents the rate of change of the infected fraction with respect to time.